UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/608,590	06/27/2003	James M. Sweet	D/A2555Q	8455

25453 7590 08/10/2005

PATENT DOCUMENTATION CENTER
XEROX CORPORATION
100 CLINTON AVE., SOUTH, XEROX SQUARE, 20TH FLOOR
ROCHESTER, NY 14644

EXAMINER	
HILLERY, NATHAN	
ART UNIT	**PAPER NUMBER**
2176	

DATE MAILED: 08/10/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/608,590	SWEET ET AL.	
	Examiner	Art Unit	
	Nathan Hillery	2176	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *29 September 2003*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-48* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-48* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *27 June 2003* is/are: a)☐ accepted or b)☒ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date 6/27/03
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

1. This action is responsive to communications: Oath or Declaration filed on 9/29/03.

2. Claims 1 – 48 are pending in the case. Claims 1, 26 and 37 are independent.

Drawings

3. New corrected drawings in compliance with 37 CFR 1.121(d) are required in this application because there are two sets of drawings submitted on the same day – one with figs 1 – 6 and another with Figs 1 – 8; consequently, the office does not know which set is correct or should be considered for examination. Applicant is advised to employ the services of a competent patent draftsperson outside the Office, as the U.S. Patent and Trademark Office no longer prepares new drawings. The corrected drawings are required in reply to the Office action to avoid abandonment of the application. The requirement for corrected drawings will not be held in abeyance.

INFORMATION ON HOW TO EFFECT DRAWING CHANGES

Replacement Drawing Sheets

Drawing changes must be made by presenting replacement sheets which incorporate the desired changes and which comply with 37 CFR 1.84. An explanation of the changes made must be presented either in the drawing amendments section, or remarks, section of the amendment paper. Each drawing sheet submitted after the filing date of an application must be labeled in the top margin as either "Replacement Sheet" or "New Sheet" pursuant to 37 CFR 1.121(d). A replacement sheet must include all of the figures appearing on the immediate prior version of the sheet, even if only one figure is being amended. The figure or figure number of the amended drawing(s) must not be labeled as "amended." If the changes to the drawing figure(s) are not accepted by the examiner, applicant will be notified of any required corrective action in the next Office action. No further drawing submission will be required, unless applicant is notified.

Identifying indicia, if provided, should include the title of the invention, inventor's name, and application number, or docket number (if any) if an application number has not been assigned to the application. If this information is provided, it must be placed on the front of each sheet and within the top margin.

Annotated Drawing Sheets

A marked-up copy of any amended drawing figure, including annotations indicating the changes made, may be submitted or required by the examiner. The annotated drawing sheet(s) must be clearly labeled as "Annotated Sheet" and must be presented in the amendment or remarks section that explains the change(s) to the drawings.

Timing of Corrections

Applicant is required to submit acceptable corrected drawings within the time period set in the Office action. See 37 CFR 1.85(a). Failure to take corrective action within the set period will result in ABANDONMENT of the application.

If corrected drawings are required in a Notice of Allowability (PTOL-37), the new drawings MUST be filed within the THREE MONTH shortened statutory period set for reply in the "Notice of Allowability." Extensions of time may NOT be obtained under the provisions of 37 CFR 1.136 for filing the corrected drawings after the mailing of a Notice of Allowability.

Double Patenting

4. The nonstatutory double patenting rejection is based on a judicially created doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the unjustified or improper timewise extension of the "right to exclude" granted by a patent and to prevent possible harassment by multiple assignees. See *In re Goodman*, 11 F.3d 1046, 29 USPQ2d 2010 (Fed. Cir. 1993); *In re Longi*, 759 F.2d 887, 225 USPQ 645 (Fed. Cir. 1985); *In re Van Ornum*, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); *In re Vogel*, 422 F.2d 438, 164 USPQ 619 (CCPA 1970);and, *In re Thorington,* 418 F.2d 528, 163 USPQ 644 (CCPA 1969).

A timely filed terminal disclaimer in compliance with 37 CFR 1.321(c) may be used to overcome an actual or provisional rejection based on a nonstatutory double patenting ground provided the conflicting application or patent is shown to be commonly owned with this application. See 37 CFR 1.130(b).

Effective January 1, 1994, a registered attorney or agent of record may sign a terminal disclaimer. A terminal disclaimer signed by the assignee must fully comply with 37 CFR 3.73(b).

5. Claims 1 – 14, 26 – 47 are provisionally rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 1 – 37 of copending Application No. 10/608587 in view of Raghavan et al. (US 6886129 B1). Raghavan et al. teach that *the present invention provides a method for identifying and enumerating groups of pages of common interest from a collection of hyper-linked pages, including the steps of: (a) identifying community cores from the collection where each core includes first and second sets of pages and each page in the first set points to every page in the second set; and (b) expanding each identified core into a full community which is a subset of the pages regarding a particular topic. To minimize the number of duplicate pages, in the hyper-links between any two pages on the same site are removed. In addition, the pages of more established sites are discarded because they might skew the results. Highly similar pages are replaced with a single page that is representative of the replaced pages, with the hyper-links previously pointing to the replaced pages now pointing to the representative page* (Column 4, lines 6 – 20), compare with **performing a document-level analysis that examines the collective set of identified candidate document pages for grouping into one or more documents**. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the invention of copending application with that of Raghavan et al. because such a combination would provide the users of copending application with *a method for identifying implicitly defined communities from a collection of hyper-linked pages* (Column 3, lines 61 – 63).

This is a provisional obviousness-type double patenting rejection.

Claim Rejections - 35 USC § 112

6. The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

7. Claims 9, 11, 33, 34, 44, and 45 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

8. Claim 9 recites the limitation "the image" in line 1. There is insufficient antecedent basis for this limitation in the claim.

9. Claim 11 recites the limitation "the document description" in line 2. There is insufficient antecedent basis for this limitation in the claim.

10. Claim 33 recites the limitation "the image" in line 1. There is insufficient antecedent basis for this limitation in the claim.

11. Claim 34 recites the limitation "the document description" in line 2. There is insufficient antecedent basis for this limitation in the claim.

12. Claim 44 recites the limitation "the image" in line 1. There is insufficient antecedent basis for this limitation in the claim.

13. Claim 45 recites the limitation "the document description" in line 2. There is insufficient antecedent basis for this limitation in the claim.

Claim Rejections - 35 USC § 103

14. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and

> the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

15. Claims 1 – 6, 10, 12 – 18, 26 – 30, 35 – 41, and 46 – 48 are rejected under 35 U.S.C. 103(a) as being unpatentable over Bharat et al. (US 6112203 A) and in further view of Raghavan et al. (US 6886129 B1).

16. **Regarding independent claim 37**, Bharat et al. teach that *we locate pages that point to at least one of the pages in the start set 201. We call this set of pages the back set 202. With the AltaVista search engine, "link:URL" queries can be used to identify back set pages for each start set page. We add one node 212 to the n-graph 211 for each page of the back set 202. Similarly, the pages pointed to by the start set 201 are located. This can be done by fetching each start set page and extracting the hyperlinks in each of the pages. The pages pointed to by the hyperlinks constitute the forward set 203. Nodes for the forward set of pages are also added to the n-graph 211. Thus, the input set of pages 204 includes the back, start, and forward sets 201-203. The input set 204 includes pages which do not directly satisfy the query, i.e., pages that do not include key words exactly as specified in the query. However, these pages may be useful because they are linked to pages of the start set. A larger n-graph 211 can be constructed by repeating this process for the back and forward sets 202-203 to add more indirectly linked pages. At this stage, the n-graph 211 has nodes 212 but no edges. After we have constructed the nodes 212, we add the directed edges 213. If a link points to a page that is represented by a node in the graph, and both pages are on different servers, then a corresponding edge 213 is added to the graph 211. Nodes representing pages on the same server are not linked. This prevents a single Web site*

with many self-referencing pages to unduly influence the outcome. This completes the n-graph 211 (Column 4, line 61 – Column 5, line 20), compare with **performing a page-level link analysis that identifies those hyperlinks on a page linking to a candidate document page further comprising a methodology of: identifying possible progression links; identifying possible table of content links, and; examining the possible progression links and the possible table of content links for common characteristics; and, performing a recursive application of the page-level link analysis to the linked candidate document page and any further nested candidate document pages thereby identified, until a collective set of identified candidate document pages is assembled**. Bharat et al. do not explicitly teach **performing a document-level analysis that examines the collective set of identified candidate document pages for grouping into one or more documents**. However, Raghavan et al. teach that *the present invention provides a method for identifying and enumerating groups of pages of common interest from a collection of hyper-linked pages, including the steps of: (a) identifying community cores from the collection where each core includes first and second sets of pages and each page in the first set points to every page in the second set; and (b) expanding each identified core into a full community which is a subset of the pages regarding a particular topic. To minimize the number of duplicate pages, in the hyper-links between any two pages on the same site are removed. In addition, the pages of more established sites are discarded because they might skew the results. Highly similar pages are replaced with a single page that is representative of the replaced pages, with the hyper-links previously pointing to the*

replaced pages now pointing to the representative page (Column 4, lines 6 – 20), compare with **performing a document-level analysis that examines the collective set of identified candidate document pages for grouping into one or more documents**. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the invention of Bharat et al. with that of Raghavan et al. because such a combination would provide the users of Bharat et al. with *a method for identifying implicitly defined communities from a collection of hyper-linked pages* (Column 3, lines 61 – 63).

17. **Regarding dependent claims 38 – 41**, Bharat et al. teach that *the nodes in the start set are first scored according to their connectivity, and the number of terms of the query that appear as unique sub-strings in the URL of the represented documents. The score is a weighted sum of the number of directed edges to and from a node and the number of unique sub-strings of the URL that match a query term* (Column 3, lines 10 – 15), compare with **the page-level link analysis includes examination of contextual clues**, **the contextual clue is a particular class of content item associated with the hyperlink**, **the class of content item is a class of text**, **the class of text is a directional word or phrase**.

18. **Regarding dependent claim 46**, Bharat et al. teach that *we assign a similarity weight to each node 213 of the sub-graph 255. Various document similarity measuring techniques have been developed in Information Retrieval to determine the goodness of fit between a "target" document and a collection of documents. These techniques typically measure a similarity score based on word frequencies in the collection and a*

target document (Column 6, lines 51 – 57), compare with **the contextual clue is the similarity of the hyperlink destination to that of other hyperlinks with the document**.

19. **Regarding dependent claim 47**, Bharat et al. teach that *we use a modified Kleinberg algorithm on the nodes of the pruned n-graph 265 to determine useful hub and authority pages. For each node of the pruned n-graph 265, we measure two scores: a hub score (HS), which estimates how good a hub the page is, and an authority score (AS), which estimates how good an authority the page is. The intuition behind our method is this: a good hub is one that points to many documents. A good authority is one that is pointed to by many documents. Transitively, an even better hub is one that points to many good authorities, and an even better authority is one that is pointed to by many good hubs* (Column 7, lines 41 – 50), compare with **the document-level analysis includes the identification of pages forming a chain of progression links**.

20. **Regarding dependent claims 18 and 48**, Bharat et al. teach that *after we have constructed the nodes 212, we add the directed edges 213. If a link points to a page that is represented by a node in the graph, and both pages are on different servers, then a corresponding edge 213 is added to the graph 211. Nodes representing pages on the same server are not linked. This prevents a single Web site with many self-referencing pages to unduly influence the outcome. This completes the n-graph 211* (Column 5, lines 13 – 20), compare with **the similarity includes the location at which**

the page is stored, and **the document-level analysis includes the identification of pages linked to by the same tables of contents**.

21. **Regarding claims 1 – 14**, the claims incorporate substantially similar subject matter as claims 37 – 47 and are rejected along the same rationale.

22. **Regarding claims 26 – 36**, the claims incorporate substantially similar subject matter as claims 37 – 47 and are rejected along the same rationale.

23. **Regarding dependent claim 15 – 17**, Bharat et al. teach that *we use do iterative connectivity analysis 310, content analysis 320, and pruning 330. This method consists of a sequence of rounds. In each round, our modified connectivity analysis is run for 10 iterations to get a listing of the (current) best hubs and authorities 315. In step 320, the pages are examined for content similarity in decreasing order of rank, alternating between the hub and the authority list. Less relevant pages are pruned* (Column 8, lines 25 – 33), compare with **the document-level analysis includes identifying the pages listed in a table of contents**, **the document-level analysis includes identifying as part of the document the page containing the table of contents**, **the document-level analysis includes the similarity of candidate pages**.

24. Claims 21 and 22 are rejected under 35 U.S.C. 103(a) as being unpatentable over Bharat et al. (US 6112203 A) and Raghavan et al. (US 6886129 B1) as applied to claims 1 – 6, 10, 12 – 18, 26 – 30, 35 – 41, and 46 – 48 above, and further in view of Huang et al. (US 6601075 B1).

25. **Regarding dependent claims 21 and 22**, neither Bharat et al. nor Raghavan et al. teach **the similarity includes similar style specifications**, and **the similarity includes similar page layout.** Huang et al. teach that *the HITS and CLEVER algorithms make use of hyperlinked structures to rank documents that share the same schema. Exemplary documents with hyperlinked structures are HTML documents. XML has given rise to a new hyperlink environment that includes documents with different schemas. In this environment, it will become increasingly important to identify high-quality schemas and documents that correctly use them. Hence, this new environment presents several previously unaddressed issues: ranking documents based on the quality of their associated schema, determining the quality of the schemas themselves, and ranking documents based on their structural properties (e.g. validity, well-formedness, etc.). The WWW today calls for a system that finds and identifies authoritative XML-documents that take these factors into account. This need, which makes use of the new dimension added by XML, has heretofore remained unsatisfied* (Column 3, lines 37 – 53), compare with **the similarity includes similar style specifications**, and **the similarity includes similar page layout**. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the combined invention of Bharat et al. and Raghavan et al. with that of Huang et al. because such a combination would allow the users of Bharat et al. and Raghavan et al. with the benefit of *an algorithm which is applied to an initial set of documents, similar to the HITS and CLEVER algorithms* (Column 3, line 66 – Column 4, line 1).

26. Claims 23 and 25 are rejected under 35 U.S.C. 103(a) as being unpatentable over Bharat et al. (US 6112203 A) and Raghavan et al. (US 6886129 B1) as applied to claims 1 – 6, 10, 12 – 18, 26 – 30, 35 – 41, and 46 – 48 above, and further in view of Law et al. (US 6754873 B1).

27. **Regarding dependent claims 23 and 25**, neither Bharat et al. nor Raghavan et al. teach **the similarity includes similar logical structure of the page content, the document-level analysis includes analysis of the topological structure of the linked pages**. Law et al. teach that *the link structure of the hyperlinked documents is analyzed in order to find hyperlinked documents that are related to and at the same level of generality of a hyperlinked document* (Column 2, lines 8 – 11), compare with **the similarity includes similar logical structure of the page content, the document-level analysis includes analysis of the topological structure of the linked pages**. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the combined invention of Bharat et al. and Raghavan et al. with that of Law et al. because such a combination would allow the users of Bharat et al. and Raghavan et al. with the benefit of *innovative techniques for finding related hyperlinked documents using link-based analysis* (Column 2, lines 6 – 8).

28. Claims 7 – 9, 11, 19, 20, 24, 31 – 34, and 42 – 45 are rejected under 35 U.S.C. 103(a) as being unpatentable over Bharat et al. (US 6112203 A) and Raghavan et al. (US 6886129 B1) as applied to claims 1 – 6, 10, 12 – 18, 26 – 30, 35 – 41, and 46 – 48 above, and further in view of Prince (US 6877002 B2).

29. **Regarding dependent claims 19, 20, 24 and 42 – 45**, neither Bharat et al. nor Raghavan et al. explicitly teach **meta-data** or **image**. However, Prince teaches that *the parsed results (from step 42 in FIG. 4) relating to the media are passed to extraction agent 68 via an extraction queue 67. Results not associated with the media are not pursued. The extraction queue 67 comprises URLs to be analyzed with respect to associated media metadata. The extraction queue 67 may comprise metadata queue entries such as media URLs, Web page URLs, Web page titles, Web page keywords, Web page descriptions, media title, media author, and media genre. Each queue entry added to the extraction queue is assigned a processing time and a priority. In an exemplary embodiment of the invention, each queue entry is given a processing time of "now" and the same default priority. The iterative seeding process increases the number of queue entries added to the extraction queue 67* (Column 7, lines 23 – 37), compare with **the similarity includes the similarity of meta-data associated with the page**, **the meta-data includes the author identification, the similarity includes the presence of at least one similar content item on each page**, **the class of content item is a class of image**, **the class of image is an image containing a directional symbol**, **a textual clue is obtained for the image**, **the contextual clue is the presence of at least one other hyperlink nearby with the document description**. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the combined invention of Bharat et al. and Raghavan et al. with that of Prince because such a combination would allow the users of Bharat et al. and Raghavan et al. with the benefit of *A method for querying metadata associated with media on a*

computer network includes separating the metadata into keywords (Column 2, lines 37 – 39).

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Nathan Hillery whose telephone number is (571) 272-4091. The examiner can normally be reached on M - F, 10:30 a.m. - 7:00 p.m.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Heather R. Herndon can be reached on (571) 272-4136. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

NH

William L. Bashore

WILLIAM BASHORE
PRIMARY EXAMINER

8/8/2005